



Col-CC: An Updated Reference Genome of *Arabidopsis thaliana*

**Xiao Dong, Raúl Y. Wijnjes, and the Community-
Consensus Arabidopsis Thaliana Reference Genome
Assembly Consortium**

**Max Planck Institute for Plant Breeding Research, Cologne
PAG 30, San Diego, 16-01-2022**

TAIR10: golden standard but incomplete

[Published: 14 December 2000](#)

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

[The Arabidopsis Genome Initiative](#)

[Nature](#) **408**, 796–815 (2000) | [Cite this article](#)

170k Accesses | **6351** Citations | **266** Altmetric | [Metrics](#)

- Accession: **Col-0**
- Methods: **Sanger** Seq and **BAC-by-BAC**
- Size: **119** Mb, missing centromeres and rDNA

Recent Col-0 assemblies: nearly complete





The genetic and epigenetic landscape of the *Arabidopsis* centromeres

MATTHEW NAISH , MICHAEL ALONGE , PIOTR WLODZIMIERZ , ANDREW J. TOCK , BRADLEY W. ABRAMSON , ANNA SCHMÜCKER, TEREZIE MANDÁKOVÁ 

BHAGYSHREE JAMGE , CHRISTOPHE LAMBING , [...], AND IAN R. HENDERSON  +14 authors [Authors Info & Affiliations](#)

SCIENCE · 12 Nov 2021 · Vol 374, Issue 6569 · DOI:10.1126/science.abi7489

High-quality *Arabidopsis thaliana* Genome Assembly with Nanopore and HiFi Long Reads






Bo Wang¹, Xiaofei Yang²  , Yanyan Jia¹, Yu Xu³, Peng Jia^{1,4}, Ningxin Dang^{1,4,5}, Songbo Wang^{1,4}, Tun Xu^{1,4}, Xixi Zhao⁵, Shenghan Gao^{1,4}, Quanbin Dong⁵
Kai Ya^{1,3,4,5}  

Molecular Plant
Correspondence

CellPress
Partner Journal

A near-complete assembly of an *Arabidopsis thaliana* genome

Pushing the limits of HiFi assemblies reveals centromere diversity between two *Arabidopsis thaliana* genomes

Fernando A. Rabanal^{1,*} , Maïke Gräff^{1,*} , Christa Lanz¹, Katrin Fritschi¹ , Victor Llaca², Michelle Lang², Pablo Carbonell-Bejerano¹, Ian Henderson³  and Detlef Weigel^{1,*} 

- Assemblies: **Col-CEN**, **Col-XJTU**, **Col-PEK**, and **Col-TUE**
- Methods: **ONT** and/or **HiFi Seq**
- Sizes: **131.6~134.6 Mb**

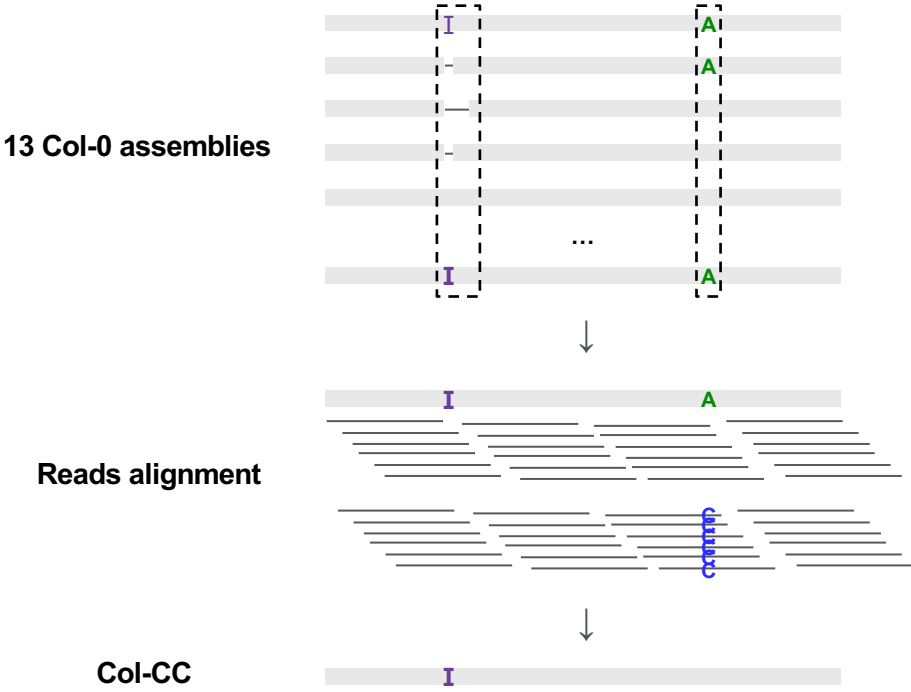
Col-CC: the community consensus assembly of Col-0



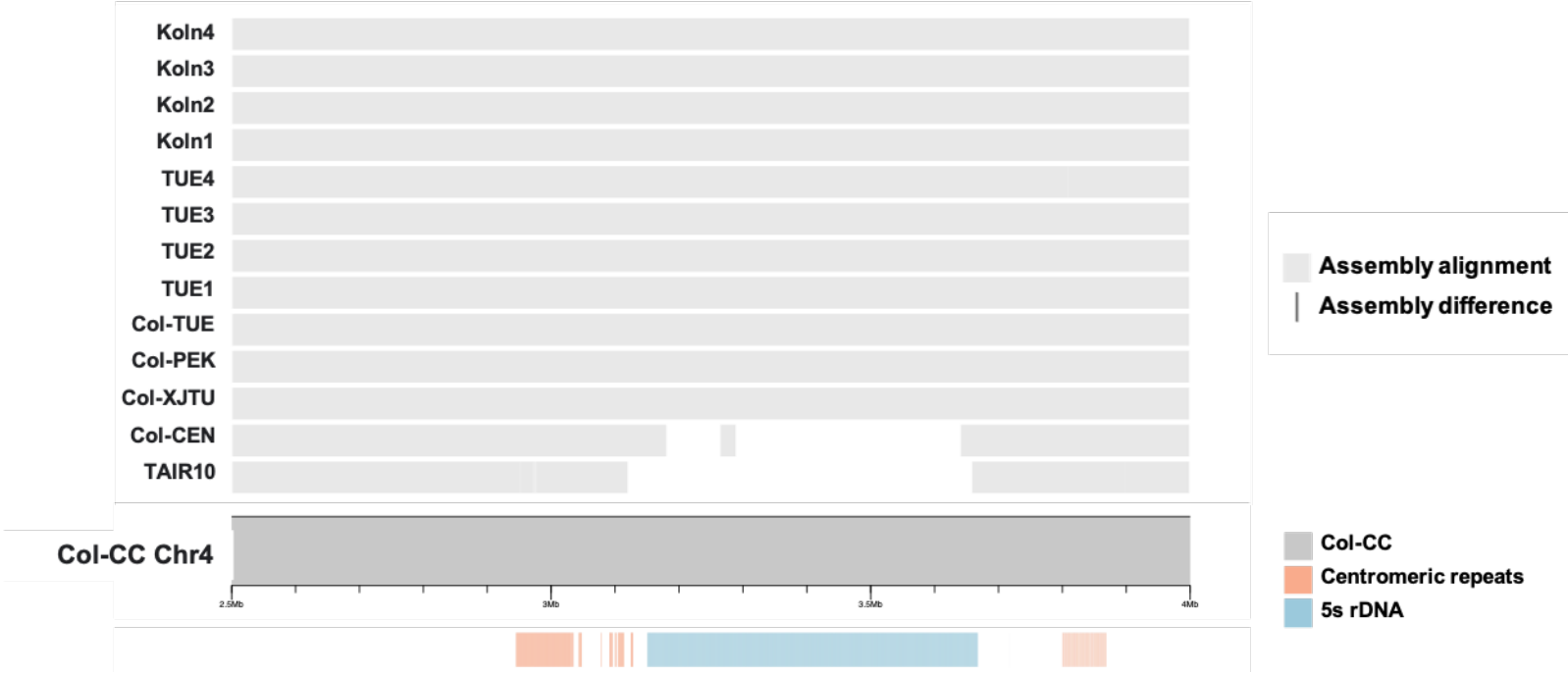
- **13 Col-0 assemblies:** TAIR10, Col-CEN, Col-XJTU, Col-PEK, Col-TUE, and 8 unpublished assemblies (TUE1-4 and Koln1-4)
- **Majority consensus** (≥ 9)

Col-CC: the community consensus assembly of Col-0

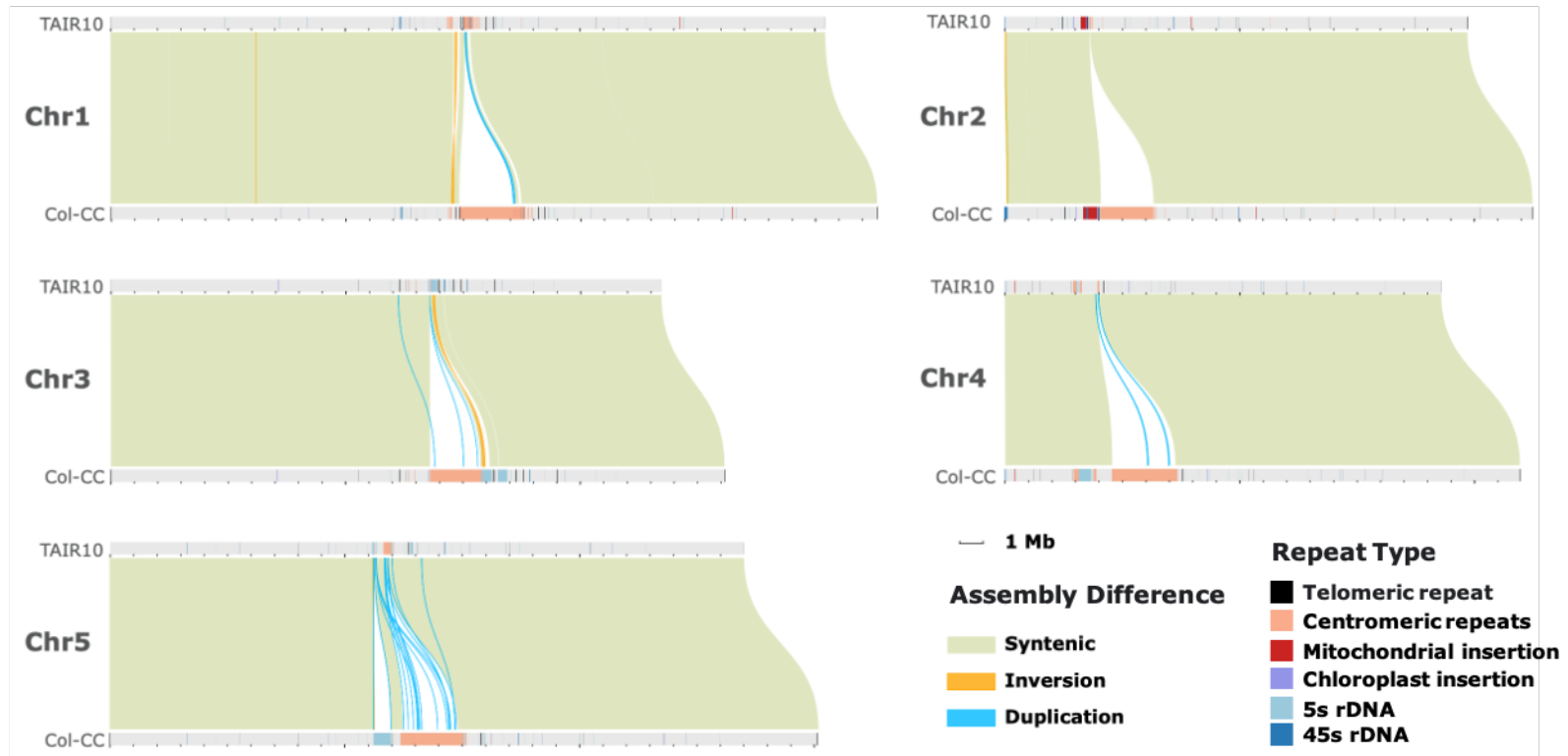
- Variant positions: 22,547
- Majority pattern: >97%
- Ambiguous cases: 663



Col-CC: resolves difficult-to-assemble regions

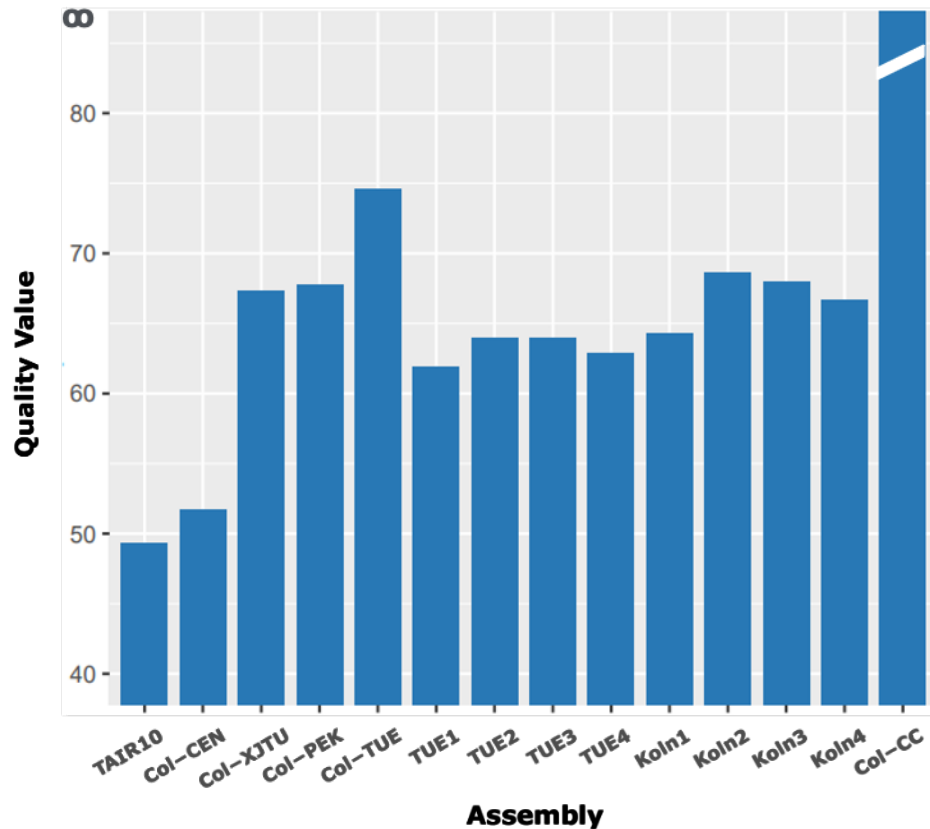


Col-CC: a complete Col-0 assembly (almost)

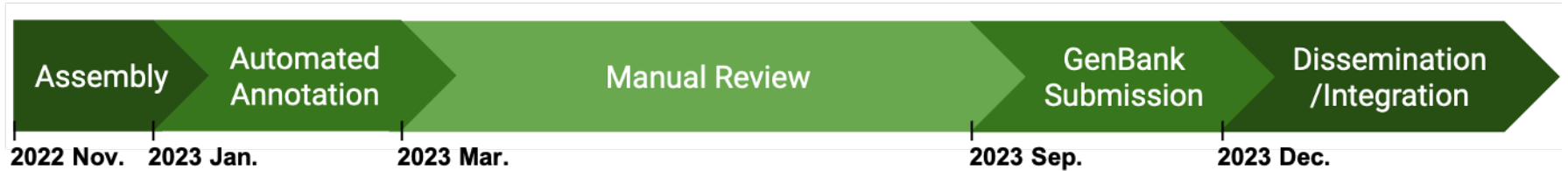


Col-CC: of insane quality

- **K-mer based evaluation**
- **Col-CC:**
 - Completeness: **99.2%**
 - QV: **77.7** (3 ambiguous base pairs)



What's next ?!



- Submitted to NCBI (BioProject **PRJNA915353**), under processing
- To undergo **NCBI** and **community-driven annotation**
- To be released as the **Arabidopsis reference genome**

Acknowledgements

Community consensus Assembly Consortium

Tanya Berardini
Nicholas Provart
Korbinian Schneeberger
Detlef Weigel
Fernando Rabanal

NCBI team

Terence Murphy

Schneeberger Lab

Raúl Y. Wijfjes

... and all other institutions and funding bodies involved

